UNIVERSITY OF RUHUNA

Faculty of Engineering

End-Semester 7 Examination in Engineering: August 2018

Module Number: EE7206 **Module Name: Machine Learning**

[Three Hours]

Index No.: .....

Instructions for Candidates:

1. This question paper consists of two parts; PART-A and PART-B carries 20 and 30 marks respectively.
2. PART-A consists of 20 questions. For PART-A, candidates should answer in the same paper (Use the ☐ space provided for answering).
3. There are 5 or 4 answers for each question. More than one ***correct*** answer or true statement may exist for one question. Candidates should mark ' ✓ ' for the ***correct*** answers and ' ✗ ' for the ***incorrect*** answers. There won't be any negative marks given. 0.2 or 0.25 marks are given for marking a ***correct*** answer as correct. 0.2 or 0.25 marks are given for marking an ***incorrect*** answer as incorrect. Unmarked answers are not given any marks.
4. PART-B consists of three essay questions. Candidates should answer them in the given answer book.

PART - A

Q1. Machine learning techniques could be effectively used for

(a) Finding the shortest path to a destination host in a communication network when the communication protocols are acting optimally ☐

(b) Forecasting the weather ☐

(c) Diagnosing a patient for dengue ☐

(d) Detecting a computer virus using a virus guard ☐

(e) Identifying a new strain of DNA which is benign to HIV where prior knowledge is inadequate ☐

Q2. The following statements are TRUE or FALSE in relation to the definition of machine learning given below regarding an Intrusion Detection System (IDS).

"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E".

(a) Observing the process of labeling a program as harmful or not is T ☐

(b) Clustering an executable program as a threat or not is T ☐

(c) Classifying an executable program as a threat or not is T ☐

(d) The probability of an executable program becoming harmful is P ☐

(e) The number of instances where programs are correctly classified as harmful or not is P ☐

Q3. The following algorithm(s) is/are used in machine learning applications.

(a) Decision trees ☐

(b) Least square regression ☐

(c) Logistic regression ☐

(d) Dijkstra's ☐

(e) Greedy ☐

Q4. Evaluate the correctness of the following statements regarding regression and classification.

(a) Predicting the next weeks average rainfall amount (in mm) is a regression problem ☐

(b) Predicting the overall weather (sunny, cloudy or rainy) for tomorrow is a regression problem ☐

(c) Selecting a particular cancer is treatable or not is a classification problem ☐

(d) Predicting the stock value (in LKR) of ABC company for the next seven days in the stock market is a classification problem ☐

(e) The same problem could be solved using both regression or classification approaches ☐

Q5. The following statements are TRUE or FALSE regarding supervised learning and unsupervised learning techniques.

(a) Developing a score predicting program for cricket matches based on player statistics and statistics at the venue of the match being held is a supervised learning problem ☐

(b) Examining the Central Processing Unit (CPU) usage of a set of programs running in a computer to group the utilization level of the programs is a unsupervised learning problem ☐

(c) A data set of heart patient statistics are examined to identify different treatment methods from grouping is a supervised learning problem ☐

(d) Clustering is an approach under supervised learning ☐

(e) Reinforcement learning is a concept defined under supervised learning ☐

Q6. The following algorithm(s) is / are used in supervised learning tasks.

(a) K-means ☐

(b) K- Nearest Neighbor (KNN) ☐

(c) Naïve Bayes ☐

(d) Support Vector Machines (SVM) ☐

(e) Neural Networks (NN) ☐

Q7. Evaluate the correctness of following statements on Reinforcement Learning (RL).

(a) RL is an area in machine learning which is inspired by behaviorist psychology ☐

(b) A person is receiving an order after time intervals of 5, 10 and 15 minutes every time he visits a restaurant, the reward is based on varying time ☐

(c) The multi-armed bandit problem captures the essence of conflict between exploration and exploitation ☐

(d) A rats reaction (turning left or right) is not affected by the probability of placing the reward (piece of cheese) at either end over successive trials in the T-maze problem ☐

(e) If a person is given a 15% discount after 2 consecutive meals in a restaurant, the reinforcement happens based on the trial ratio rather than the trail times ☐

Q8. Evaluate the correctness of following statements regarding Support Vector Machines (SVM).

(a) Support vectors could be determined by the value of the parameter α ☐

(b) SVM uses an optimization approach rather than a greedy search for solution which improves its efficiency ☐

(c) Support vectors are the elements of the training set that would change the position of the decision hyperplane if removed ☐

(d) Lagrangian duality is used for maximizing the margin of the decision hyperplane in SVM ☐

(e) Gaussian kernel is better suited for non-linear classifications ☐

Q9. Mark the TRUE and FALSE statements regarding probabilistic classification schemes.

(a) Naïve Bayes assumption is $P\left(o_1, o_2, \ldots, o_n \middle| h_j\right) = \sum_i P\left(o_i \middle| h_j\right)$ ☐

(b) Bayes classifier only relies on the prior probabilities for classification ☐

(c) Meta classification is a significant property in Bayes classifier ☐

(d) Bayes classifier computes the maximum a posteriori (MAP) hypothesis ☐

(e) Probability based classifiers perform better with predictive problems ☐

Q10. Evaluate the correctness of following statements regarding feature extraction mechanisms.

(a) $\hat{x} = \dfrac{\sqrt{x/x_{\min}} - \sqrt{x_{\min}/x_{\max}}}{x_{\max} - x_{\min}}$ would be a good normalization model. ☐

(b) A good feature should produce high correlation within the class while highly correlated features should not be redundant. ☐

(c) A feature is always a numerical value which is measured from a sensor. ☐

(d) Features could be invented either automatically or manually. ☐

(e) Circumference is a good feature for classifying white blood cells. ☐

Q11. Which of the following statement(s) is/are TRUE or FALSE ?

(a) Suppose you have a multi-class classification problem with three classes, trained with a 3 layer network. Let $a_1^{(3)} = (h_\Theta(x))_1$ be the activation of the first output unit, and similarly $a_2^{(3)} = (h_\Theta(x))_2$ and $a_3^{(3)} = (h_\Theta(x))_3$. Then for any input x, it must be the case that $a_1^{(3)} + a_2^{(3)} + a_3^{(3)} = 1$. ☐

(b) A two-layer (one input layer, one output layer; no hidden layer) neural network can represent the OR function. ☐

(c) Any logical function over binary-valued (0 or 1) inputs x1 and x2 can be (approximately) represented using some neural network. ☐

(d) The activation values of the hidden units in a neural network, with the sigmoid activation function applied to every layer, are always in the range (0, 1). ☐

Q12. Consider the following neural network which takes two binary-valued inputs x1, x2∈{0,1} and outputs hΘ(x). Which following logical functions compute (approximately)

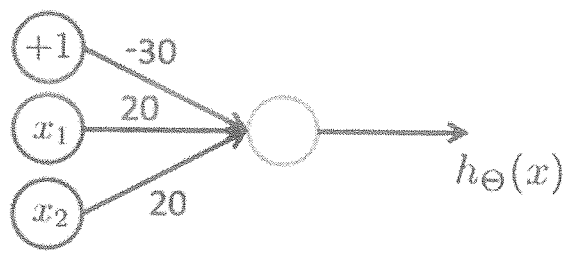

Figure Q12

(a) NAND (meaning "NOT AND") ☐

(b) AND ☐

(c) OR ☐

(d) XOR (exclusive OR) ☐

Q13. Consider the neural network given below. The following equation(s) correctly compute(s) the activation $a_1^{(3)}$. (Note: $g(z)$ is the sigmoid activation function)

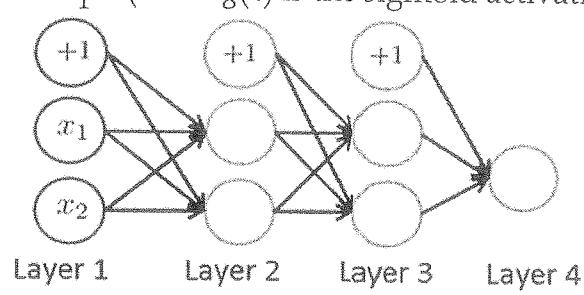

Figure Q13

(a) $a_1^{(3)} = g\left(\Theta_{1,0}^{(2)} a_0^{(2)} + \Theta_{1,1}^{(2)} a_1^{(2)} + \Theta_{1,2}^{(2)} a_2^{(2)}\right)$ ☐

(b) $a_1^{(3)} = g\left(\Theta_{1,0}^{(2)} a_0^{(1)} + \Theta_{1,1}^{(2)} a_1^{(1)} + \Theta_{1,2}^{(2)} a_2^{(1)}\right)$ ☐

(c) $a_1^{(3)} = g\left(\Theta_{1,0}^{(1)} a_0^{(2)} + \Theta_{1,1}^{(1)} a_1^{(2)} + \Theta_{1,2}^{(1)} a_2^{(2)}\right)$ ☐

(d) $a_1^{(3)} = g\left(\Theta_{2,0}^{(2)} a_0^{(2)} + \Theta_{2,1}^{(2)} a_1^{(2)} + \Theta_{2,2}^{(2)} a_2^{(2)}\right)$ ☐

Q14. A neuron with 4 inputs has the weight vector w = [1; 2; 3; 4]T and a bias μ = 0 (zero). The activation function is linear, where the constant of proportionality equals 2, that is, the activation function is given by f(net) = 2 £ net. If the input vector is x = [4; 8; 5; 6]T then the output of the neuron will be

(a) 1 ☐
(b) 56 ☐
(c) 59 ☐
(d) 112 ☐
(e) 118 ☐

Q15. 5. You are using the neural network pictured below and have learned the parameters $\Theta^{(1)}=\begin{bmatrix}1 & -1.5 & 3.7\\ 1 & 5.1 & 2.3\end{bmatrix}$ (used to compute $a^{(2)}$) and $\Theta^{(2)}=[1 \quad 0.6 \quad -0.8]$ (used to compute $a^{(3)}$) as a function of $a^{(2)}$). Suppose you swap the parameters for the first hidden layer between its two units so $\Theta^{(1)}=\begin{bmatrix}1 & 5.1 & 2.3\\ 1 & -1.5 & 3.7\end{bmatrix}$ and also swap the output layer so $\Theta^{(2)}= [1 \quad -0.8 \quad 0.6]$ value of the output $h_\Theta(x)$

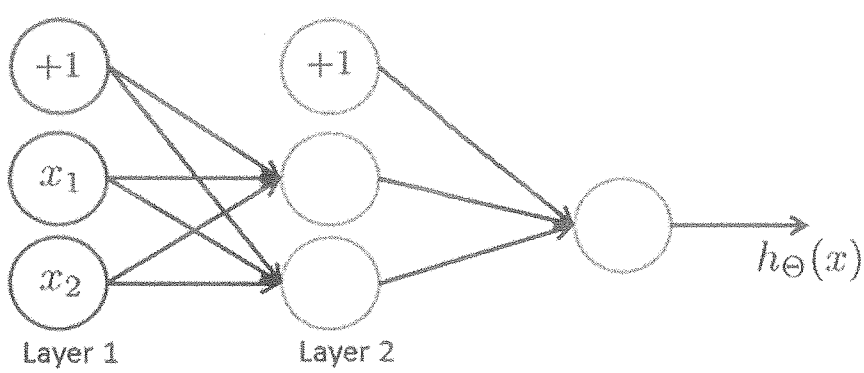

Figure Q15

(a) will stay the same. ☐

(b) will increase. ☐

(c) will decrease. ☐

(d) may increase or decrease (Insufficient information to tell). ☐

Q16. The following statement(s) describe(s) the reproduction process.

(a) Randomly change a small part of some strings. ☐
(b) Randomly change the fitness function. ☐
(c) Randomly combine the genetic information from 2 strings. ☐
(d) Randomly generate small initial values for the weights. ☐
(e) Randomly pick strings to make the next generation. ☐

Q17. The following statement(s) describe(s) the mutation process.

(a) Randomly change a small part of some strings. ☐
(b) Randomly change the fitness function. ☐
(c) Randomly combine the genetic information from 2 strings. ☐
(d) Randomly generate small initial values for the weights. ☐
(e) Randomly pick strings to make the next generation. ☐

Q18. If crossover between chromosomes in search space does not produce significantly different offspring, what does it imply? (if offspring consist of one half of each parent)

(i) The crossover operation is not successful.
(ii) The solution is about to be reached.
(iii) Diversity is poor that the parents involved in the crossover operation are similar.
(iv) The search space of the problem is not ideal for GAs to operate.

The correct and incorrect statement(s) is/are,

(a) ii, iii & iv only ☐
(b) ii & iii only ☐
(c) i, iii & iv only ☐
(d) all of the above ☐

c) Consider the following data coordinates in a Voronoi diagram to be clustered using k-means clustering algorithm.
(0.5, 4.5) , (1, 1) , (0.5, 2) , (2, 1) , (4, 1) , (6, 1) , (3, 2.5) , (5, 2) , (1, 3) , (2, 4.5) , (1.5, 5.5) , (1, 6.5)
The initial centroids for k = 3 are $\mu_1 \equiv (3,7), \mu_2 \equiv (4,4)$ and $\mu_3 \equiv (7,1)$ respectively.

i) Illustrate the Voronoi diagram by assigning and representing the clusters using the Euclidian distance to the centroids.

ii) Converge the data set through two clustering phases for determining the finalized cluster centroids.

iii) Classify the new data instances (2, 3) and (4.5, 2).

[4 Marks]

Q2. a) Identifying spam emails is a major function carried out by modern day email servers. In order to identify an email as a spam email, several factors could be considered. The spam filter could be configured to detect words from the vocabulary w, which include the words such as {get rich, buy, invest, win, bonanza...etc.}. Table Q2 shows the several instances being classified as spam or valid emails with corresponding features.

Table Q2

	No sender (F1)	No subject (F2)	No text in the email body (F3)	Include words from w (F4)	Contain executable attachments (F5)	Email Detection (D)
1	Y	Y	Y	Y	Y	S
2	Y	Y	Y	N	Y	S
3	N	Y	Y	Y	N	S
4	Y	N	Y	Y	N	S
5	Y	N	Y	Y	Y	S
6	N	Y	N	Y	Y	S
7	Y	N	N	Y	Y	S
8	N	Y	Y	N	Y	S
9	N	N	N	N	N	V
10	N	Y	N	N	N	V
11	N	N	N	N	Y	V
12	N	N	N	N	N	V
13	N	N	Y	Y	Y	V
14	N	Y	N	N	Y	V
15	N	Y	N	Y	N	V
16	N	N	Y	Y	N	V

i) Starting from the Bayes theorem, derive the Naïve Bayes classifier while explaining the terms given in the below equation.

$$P(h|o)=\frac{P(o|h)P(h)}{P(o)}$$

ii) Estimate the prior probabilities and compute the conditional probabilities for the Naïve Bayes classification explained above.

iii) Use the Naïve Bayesian classifier to classify the feature vectors V1 = (Y, Y, Y, N, Y) and V2 = (N, N, N, Y, Y).

[5 Marks]

b) SVM uses the Lagrange multipliers to optimize the boundary/ street width of $\frac{2}{\|\vec{w}\|}$. If the primal function is given by

$$L(\vec{w},b)=\frac{1}{2}\vec{w}.\vec{w}-\sum_{i=0}^{N-1}\alpha_i\left[t_i\left((\vec{w}.\vec{x}_i)+b\right)-1\right]$$

i) Construct the dual function from the primal function.

ii) Determine the marginal hyperplanes of the section Q1. b) , if the size of the margin is 3.

iii) Are there any support vectors in the data set given in section Q1. b) ?

[2.5 Marks]

c) Mention three machine learning applications and use your imagination to formulate the operation of one such application. Explain the proposing classification or clustering mechanism descriptively. (Use diagrams when explaining)

[2.5 Marks]

Q3. a) Genetic Algorithms are based on Darwin's theory of evolution and the survival of the fittest. Keywords are population, individuals, fitness, selection and genetic operators (crossover and mutation). Explain the genetic algorithm.

[3 Marks]

b) Crossover is one of the most important genetic operators of the Genetic Algorithm. From the traveling salesperson problem (TSP), you are given a list of cities to visit as a potential solution. Given the two parents *A* and *B*, create the resulting children *C* and *D* by applying different cross-over operators that keep the chromosome valid (I.e. each city may only occur once in the chromosome). Create child *C* by applying for instance Partially Mapped Xover (PMX) on parent *A* and *B*, and create child *D* by applying Order Crossover on parent *A* and *B*. The cross-over points are indicated on parent *A* and *B*. Apply inversion and insertion mutation to child C and D respectively.

A: 3 9 7 | 4 2 1 | 5 6

B: 1 8 7 | 4 3 5 | 6 2

[4 Marks]

c) Using an example, explain why it is important to have a mutation operator in a genetic algorithm.

[3 Marks]